



University of Groningen

Grid computing en e-science

Petkov, Nikolai

Published in:
EPRINTS-BOOK-TITLE

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2003

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Petkov, N. (2003). Grid computing en e-science. In EPRINTS-BOOK-TITLE University of Groningen, Johann Bernoulli Institute for Mathematics and Computer Science.

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Onderzoek

3.1 Grid computing en e-science

Nikolai Petkov*

Inleiding

Het is al geruime tijd mogelijk om met het *file transfer protocol (ftp)* of *hyper text transfer protocol (http)* databestanden van computers op afstand via het wereldwijde computernetwerk op te halen. Met *telnet* en *remote procedure calls (rpc)* kan men werksessies op computers op afstand starten respectievelijk rechtstreeks commando's laten uitvoeren. Op soortgelijke wijze houdt het overgrote deel van de huidige internet-technologieën zich bezig met het uitwisselen van gegevens tussen computers. Gecoördineerd gebruik van computers die zich op verschillende geografische locaties of administratieve netwerkdomeinen bevinden, komt nog zelden voor. Het ogenschijnlijke gemak echter, waarmee men multimediale informatie via het World Wide Web kan verkrijgen, de opkomst van leveranciers van reken-, opslag- en toepassingsdiensten en, niet in de laatste plaats, de populaire toepassingen voor het uitwisselen van muziek en films en de on-linespelletjes over het internet hebben onderzoekers op de gedachte gebracht dat het wereldwijde computernetwerk méér voor mensen en organisaties kan betekenen dan nu het geval is. Dit is het doel van gridtechnologieën. Het nieuwe van gridtechnologie is de integratie, het gecoördineerd gebruik en het delen van ICT-bronnen (computers, programmatuur en gegevens) die zich op één of op verschillende locaties bevinden. Webstandaarden en -technologieën hebben een universele transparante toegang tot documenten mogelijk gemaakt; gridtechnologieën zouden dezelfde betekenis kunnen hebben voor computerfaciliteiten, -gegevens en -applicaties.

Wat is een grid?

De term (*computer*)*grid* werd medio jaren negentig gelanceerd als een concept voor gedistribueerde computerinfrastructuur die gecoördineerd gebruikt kan worden. De naam is door de analogie met het *electrical grid* geïnspireerd (het Amerikaanse elektriciteitsnetwerk wordt *the grid* genoemd), met dien verstande dat een computergrid computerdiensten levert in plaats van elektriciteit. ICT-bronnen, zoals reken-

* Voor commentaar op dit hoofdstuk en aanvullende informatie zegt de auteur dank aan: dr. P. Aerts, J. Bokma, prof.dr. H. Butcher, H. Paas en dr. L. Plugge.

capaciteit, data-opslag, applicatiesoftware en gegevens, worden geïntegreerd via het netwerk aangeboden als een elektronische nutsvoorziening (*e-utility*). ICT-diensten zouden voor de gebruiker even gemakkelijk 'uit de muur' of 'uit het stopcontact' worden geleverd als elektriciteit.

De proponenten van computergrids zien een zekere overeenkomst tussen de ontwikkeling van ICT en die van het elektriciteitsnetwerk. In de beginjaren van de elektriciteit werden generatoren voor lokaal gebruik gebouwd. Vervolgens is uit de praktijk gebleken dat het voordeliger en efficiënter is afnemers en leveranciers van elektriciteit in een groot netwerk te verbinden. Een fabriek die op bepaalde momenten met pieken in de belasting wordt geconfronteerd, hoeft, om in de eigen behoefte te voorzien, geen grote elektriciteitscentrale te bouwen die in de rest van de tijd onderbenut wordt. De ICT-infrastructuur van organisaties wordt echter nog steeds door de pieken in de behoefte bepaald, met als gevolg een overcapaciteit in de rest van de tijd en een gemiddelde lage efficiëntie.

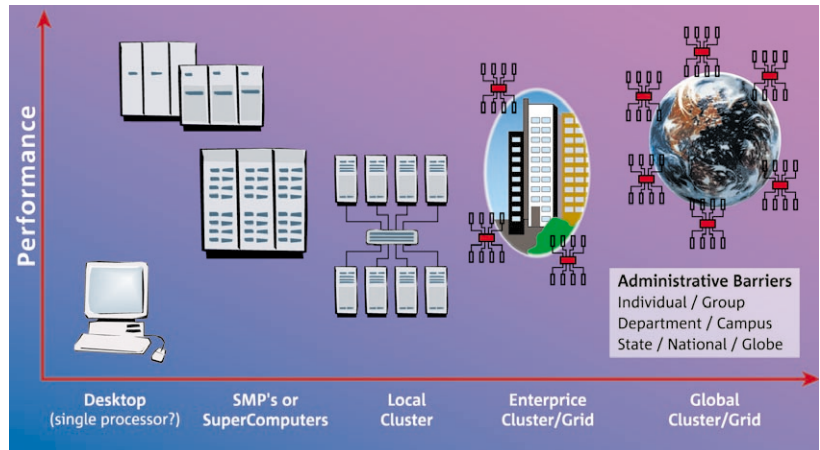
Grids gedefinieerd

Grids zijn dynamische gedistribueerde ICT-omgevingen die softwaretoepassingen in staat stellen ICT-bronnen, zoals reken- en opslagfaciliteiten, gegevensbanken, displays en instrumenten, via een computernetwerk te integreren. De integratie kan binnen een afdeling, een instituut of een organisatie plaatsvinden of de grenzen van organisaties en netwerkdomeinen overstijgen. De gedeelde ICT-bronnen kunnen zich op dezelfde of op verschillende geografische locaties bevinden. Voor de gebruiker ziet een grid er uit als een integraal computersysteem met alle benodigde faciliteiten. De gebruiker merkt niet waar deze faciliteiten zich bevinden.

De term grid lijkt inmiddels gerelateerde begrippen, zoals *internet-computing*, *network-computing*, *e-science*, *virtual laboratories*, *e-business*, *computing on demand* en *next-generation data center*, te vervangen, te verdringen of in zich op te nemen. Inhoudelijk is deze begrippenconvergentie overigens niet helemaal juist, omdat de term grid voor een *enabling* infrastructuur staat, terwijl de andere begrippen een bepaald soort gebruik van het internet benadrukken, waarvoor grids deels een voorwaarde zijn.

MIDDLEWARE ICT-bronnen worden geïntegreerd door middel van specifieke *middleware*. Deze middleware zorgt ervoor dat aan de behoeften aan rekenkracht en opslagcapaciteit wordt voldaan, zonder dat een gebruiker iets van het gedistribueerde karakter van het achterliggende systeem merkt. De grid-middleware laat de faciliteiten, die op verschillende locaties gevestigd kunnen zijn, voor de gebruiker verschijnen als een computer op zijn eigen kamer, maar zonder capaciteitsbeperkingen. Door grid-middleware worden gebruikersbehoeften en beschikbare faciliteiten bij elkaar gebracht. De middleware vervult onder andere de rol van *resource broker*: hij bewaakt de toestand van het netwerk en de faciliteiten en zorgt voor een optimale toewijzing van netwerk-, reken- en opslagcapaciteit tegen de voordeligste prijs. Middleware maakt het mogelijk ook andere dan ICT-apparatuur, zoals specifieke meetinstrumenten, op afstand (inbraak-)veilig te gebruiken.

Figuur 34
De evolutie van
computerplatforms



NETWERK-INFRASTRUCTUUR Om een grid *binnen* een organisatie (organisatiegrid: zie 3.1.3.6) te realiseren, zijn de huidige bandbreedte en de bandbreedte die voor de komende jaren in Nederland is gepland, voldoende. Men hoeft dus niet op de volgende netwerk-upgrade te wachten.

Het grootschalig toepassen van grids, waarbij meerdere organisaties in meerdere netwerk-domeinen zijn betrokken, kan echter tot meer bulk-datatransport gaan leiden. Daardoor is misschien per applicatie niet meer bandbreedte nodig, maar voor de som van de applicaties al snel wel. Daarnaast zullen er altijd specifieke toepassingen zijn waarbij de beschikbare bandbreedte achterloopt bij de wensen van de gebruikers. De verwerking van de enorme datastromen die door de meetapparatuur in deeltjesversnellers of door *arrays* van radiotelescopen gegenereerd worden, is een typisch voorbeeld hiervan. Het *Particle Physics Data Grid* bijvoorbeeld zal gebaseerd zijn op een communicatienetwerk met hoge snelheid en hoge capaciteit, dat in de komende drie jaar duizend keer sneller dan de huidige systemen zal worden.

De ervaringen met gridtoepassingen laten zien dat in veel gevallen niet het landelijke netwerk, maar juist de universitaire *backbone* of de aansluiting daarop de bottleneck is. Terwijl men op landelijk en afdelingsniveau over bandbreedte op Gbp/s-niveau beschikt, wordt de uiteindelijke bandbreedte door het universitaire netwerk en in het bijzonder door de hoge prijs van overgangapparatuur met een factor tien beperkt. Gegarandeerde bandbreedte op alle niveaus kan voor veel organisatie-overstijgende gridtoepassingen essentieel zijn.

BELEID EN ORGANISATIE Grids realiseren is niet alleen een kwestie van technologie. In veel gevallen, zoals bij een klein grid dat in een organisatie-onderdeel op basis van intranet wordt gerealiseerd, is de gridtechnologie uit de experimenteerfase gegroeid en is zij rijp voor inzet in de praktijk. Beleid, organisatie en management zijn nodig om locale faciliteiten dynamisch aan een grid toe te wijzen. Om bijvoorbeeld de Windows-pc's van de verschillende secretariaten en onderwijs-pc-pools van een organisatie voor gridtoepassingen te kunnen inzetten, moet langs strakke organisatorische lijnen en onder een centrale regie aan diverse voorwaarden worden voldaan. Op afdelingsniveau kan men met veiligheidsbedenkingen van systeembeheerders te maken krijgen. Op facultair en universitair niveau leidt het gebruik van gedeelde ICT-bronnen via een

grid tot verwevenheid van de ICT-budgetten van verschillende onderdelen van de organisatie.

VOORDELEN Tot de algemene voordelen van grids horen onder meer: reductie van de totale kosten van eigendom; hogere efficiëntie van het gebruik van ICT-bronnen door schaalvergroting; en verlaging van de drempels voor de dynamische integratie van organisaties bij fusies of uitvoering van gezamenlijke projecten.

In de context van hoger onderwijs en onderzoek kunnen grids onder meer worden gebruikt voor:

- een efficiënter gebruik van de aanwezige ICT-infrastructuur;
- reductie van de directe ICT-investeringen en beheerskosten;
- toegang tot grote computercapaciteit door integratie van gedistribueerde rekenfaciliteiten;
- gebruik van geïntegreerde gedistribueerde gegevensbanken en in programma's vertaalde toepassingen;
- interactief samenwerken (tussen wetenschappers en wetenschappelijke groepen internationaal) via het computernetwerk, en experimenten uitvoeren met apparatuur op afstand.

DE ROL VAN EN VOOR BEDRIJVEN In de beginjaren lag het initiatief voor het ontwikkelen van de grids in handen van door de overheid gefinancierde wetenschappelijke instellingen. De activiteiten richtten zich vooral op het ontwikkelen van specifieke wetenschappelijke applicaties voor onderzoeksgebieden als hoge-energiefysica en radioastronomie.

De gridtechnologie heeft inmiddels de academische couveuse verlaten: ze wordt door bedrijven niet alleen toegepast maar zelfs als product of dienst aangeboden. Deze technologie maakt een nieuw business-model mogelijk – namelijk het omgaan met ICT-faciliteiten zoals met een nutsvoorziening – dat voor het bedrijfsleven zeer aantrekkelijk is. Fusies, acquisities of interne herstructureringen maken het vaak noodzakelijk, dat medewerkers en teams die zich op verschillende locaties bevinden toch nauw samenwerken. In zulke situaties kan de gridtechnologie ervoor zorgen dat de gedistribueerde ICT-infrastructuur geen belemmering vormt voor de bedrijfsprocessen.

Het belang van gridtechnologie wordt momenteel in voldoende mate in het bedrijfsleven onderkend (tenminste door grote bedrijven). De belangstelling ervoor is groot. Intel zet bijvoorbeeld gridtechnologie in om aan voldoende rekenkracht te komen voor het ontwikkelen van nieuwe chips. Het farmaciebedrijf Pfizer maakt gebruik van gridtechnologie om gedistribueerde gegevensbanken, die nodig zijn bij de ontwikkeling van nieuwe geneesmiddelen, te integreren en beschikbaar te stellen aan onderzoeksteams die op verschillende locaties werken. Grids worden gebruikt bij de analyse van financiële risico's en in de olie-industrie. SUN Microsystems heeft een eigen grid met meer dan 7.500 processors, die verspreid zijn over drie staten van de VS. Volgens SUN wordt 98% efficiency bereikt bij het gebruik van dit grid. Zelfs in de industrie voor computerspelletjes zijn er plannen om gridtechnologie in nieuwe producten toe te passen. Zo hebben Sony, Toshiba en IBM de gezamenlijk te ontwikkelen chipset voor de Playstation3 'grid' genoemd. De naam suggereert dat men ook op dit terrein gebruik wil maken van gedistribueerde verwerking en het delen van software, verwerkingskracht en gegevens op spelcomputers die via het internet verbonden zijn.

De activiteiten van diverse grote ICT-bedrijven, zoals IBM, Hewlett-Packard, Oracle en SUN Microsystems, op het terrein van ontwikkeling, ondersteuning en gebruik van gridtechnologie zijn de laatste jaren enorm toegenomen in intensiteit en focus. Ze zullen beslist leiden tot een versnelde ontwikkeling van gridtechnologie, tot schaalvergroting van de gridactiviteiten en uiteindelijk tot uitbreiding naar veel commerciële toepassingen en brede acceptatie van het *e-utility*-bedrijfsmodel.

Categorieën van grids

Grids kunnen op basis van het type ICT-bron dat wordt gedeeld of het doel dat wordt nagestreefd, worden gecategoriseerd in computerfaciliteiten-, reken-, data-, apparatuur- en toepassingsgrids. Naargelang de schaal waarop de integratie en het delen van ICT-bronnen plaatsvindt, kan men spreken van organisatie-, partner- en servicegrids.

COMPUTERFACILITEITEN-GRIDS Bij een computerfaciliteitengrid gaat het in eerste instantie om dynamische integratie van reken- en opslagfaciliteiten. De integratie vindt via het computernetwerk plaats en kan tot een afdeling, instituut, faculteit of universiteit zijn beperkt of de grenzen van de organisatie overstijgen.

REKENGRIDS Als veel rekenkracht door het gecoördineerd gebruik van veel computers het doel is, spreekt men ook van een rekengrid (*compute-grid*). Hieraan ligt de gedachte ten grondslag die ook de basis is van het parallel rekenen: een toepassing wordt gesplitst in onderdelen die simultaan op verschillende computers kunnen worden uitgevoerd. Meestal gaat het hierbij om wetenschappelijke en technische simulaties waarmee veel berekeningen gemoeid zijn.

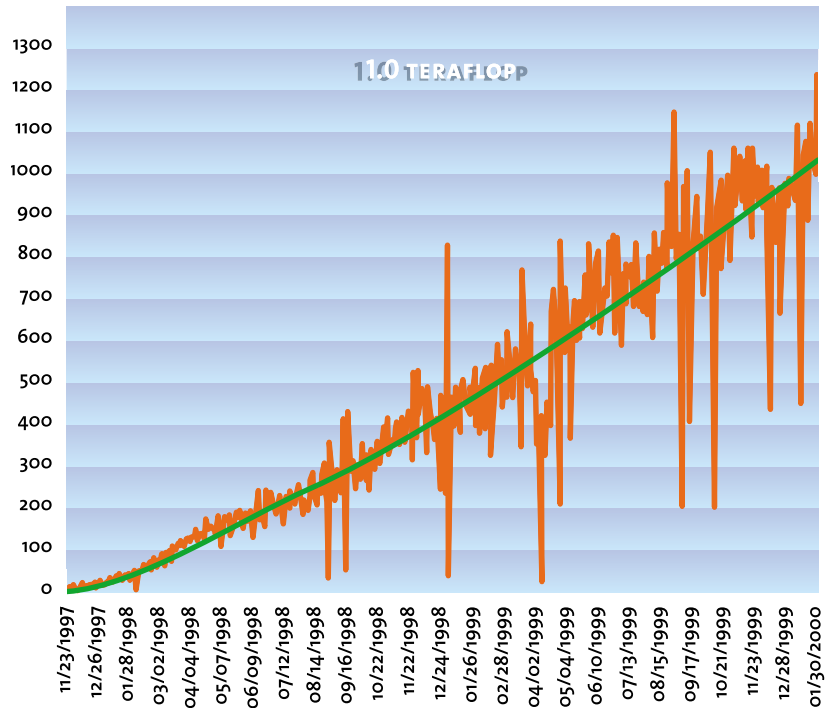
Het idee van *grid-computing* is uit de behoefte aan rekenkracht in het academisch onderzoek ontstaan. Bij verschillende Amerikaanse centra voor *super-computing* wordt hiermee geëxperimenteerd, met de bedoeling méér rekenkracht en data-opslagcapaciteit beschikbaar te stellen aan grootschalige wetenschappelijke projecten dan door een enkele supercomputer geboden kan worden. Volgens dit concept krijgen deelnemers aan een wereldwijd computergrid toegang tot praktisch onbeperkte rekenkracht en opslagvermogen.

Om een idee te geven van de mogelijkheden van de meest geavanceerde supercomputers, noemen we de Earth Simulator in Japan: zijn rekenkracht bedraagt op dit moment 35 teraflops (35.000.000.000.000 *floating-point*-operaties per seconde). Dit komt overeen met de cumulatieve rekenkracht van circa 30.000 krachtige pc's. De rest van de top-5, alle vier in de VS, hebben een rekenkracht tussen 5,6 en 7,7 teraflops.

De drijfveer achter de ontwikkeling van rekengrids is de permanent stijgende behoefte aan verwerkingskracht. Zodra wetenschappers een nieuwe krachtige computer tot hun beschikking krijgen die aan hun huidige behoeften voldoet, komen ze met vraagstellingen die nog krachtiger computers vereisen. Computerinfrastructuur-oplossingen die de grenzen van een organisatie overschrijden, vormen een voor de hand liggende aanpak van dit probleem. Dit is in het bijzonder het geval als men met toepassingen te maken heeft, die pieken in de belasting veroorzaken.

De technologie voor het bouwen van rekengrids is deels ontstaan in samenhang met

Figuur 35
Performance van het
Entropia-netwerk voor
het bepalen van het
grootste Mersenne
priemgetal



intranet binnen organisaties. Een organisatie heeft veel computers (servers, werkstations, desktop-pc's) die het overgrote deel van de tijd niet worden gebruikt. Als er behoefte is aan rekencapaciteit, kunnen deze computers aan elkaar worden gekoppeld (bijvoorbeeld 's nachts) om als één parallelle computer de deelberekeningen van een toepassing uit te voeren. Er zijn volwassen commerciële softwareproducten, meestal server-georiënteerd, die een dergelijke integratie van faciliteiten binnen een intranet mogelijk maken.

De meeste huidige gridonderzoeksprojecten en commerciële producten zijn op rekengrids gericht. Rekengrids op basis van werkstations en pc's bieden een voordelige oplossing voor veel problemen die met grootschalig rekenen te maken hebben. Ze zijn echter zeker geen oplossing voor alle rekenintensieve problemen en moeten dan ook niet als de ultieme vervanger van supercomputers worden gezien.

Rekenproblemen die volledig te splitsen zijn in relatief grote onafhankelijke deelproblemen, kunnen op meerdere computers tegelijk worden aangepakt. Een voorbeeld uit de cryptografie is het ontbinden van grote getallen in priemfactoren. Maar rekenproblemen waarbij op gezette tijden intensieve communicatie tussen de processen plaatsvindt, kunnen weliswaar op parallelle computers worden aangepakt, maar lenen zich niet voor parallelle verwerking via een grid.

De alles beperkende factor is de lichtsnelheid. Ongeacht de bandbreedte van het netwerk is de latentie bij berichtuitwisseling, zelfs met weglaten van alle protocol-tijdverliezen en het feit dat de lichtsnelheid in glas nog lager is dan in vacuüm, een harde beperkende factor. Een grid dat een computer in Groningen en één in Delft omvat, en waarop een parallelle job draait, zal bij elke uitwisseling van een bericht

aan latentietijd zoveel rekencycli verliezen, dat het in het geval van frequente data-uitwisseling nooit een alternatief voor een supercomputer kan zijn. In ieder geval is het fout te denken, dat de installatie van een grid dat een groot aantal computers in Nederland verbindt een nationale supercomputer voor grootschalig rekenwerk overbodig zou maken.

DATAGRIDS Naast het delen van rekenfaciliteiten is er behoefte om databronnen te delen. In de academische wereld zijn voorbeelden hiervan te vinden op de terreinen van hoge-energiefysica, sterrenkunde, scheikunde, genetica, bio-informatica, mens- en sociale wetenschappen en de kunsten. Ook bedrijven, zoals banken, verzekeraars, olie- en luchtvaartmaatschappijen en de farmaceutische industrie, werken met grote data-systemen die vanuit verschillende locaties door meerdere gebruikers geraadpleegd moeten worden. In veel gevallen gebeurt het delen van gegevens nog steeds door overdracht van bestanden met file transfer protocol (ftp of scp, rcp en aanverwante kopieer-programma's), waarbij de gebruiker een sessie moet initiëren en moet aangeven welke bestanden uit welke subdirectory's van een systeem naar welke subdirectory's in een ander systeem gekopieerd dienen te worden. Door te werken met scripts kan dit proces deels worden geautomatiseerd. Dit leidt echter onvermijdelijk tot complexe informatie-infrastructuren die op bestandenreplica's zijn gebaseerd, zonder garanties dat de gegevens actueel zijn op het moment dat ze worden gebruikt. Bovendien ontstaat door deze aanpak onnodig veel netwerkverkeer. Tot op heden is voor *data-sharing* tussen machines voornamelijk het *NFS-protocol* in gebruik. NFS heeft een aantal problemen. Vooral door de gebruikte netwerk-technologie (over UDP en dus niet gegarandeerd foutloos) is zij alleen goed te gebruiken op een intranet. Verder zijn er problemen met NFS op het terrein van beveiliging en authenticatie. Zo ontbreken een adequate dataversleuteling, translatie van de gebruikersidentificatie en overige beveiligingen van de dataconnectie. Er zijn wel systemen en protocollen voor data-sharing die dit beter geregeld hebben, zoals *Kerberos*, *DFS (Distributed File System)* en *AFS (Andrew File System)*. Een wijdverbreide standaardisatie ontbreekt echter tot op dit moment. Reden te meer dit probleem in het kader van grid-computing aan te pakken.

Bij een datagrid, ook wel *informatiegrid* genoemd, gaat het om toegang tot informatiebronnen die op verschillende computers, administratieve domeinen of geografische locaties gevestigd zijn. De drijvende kracht achter de ontwikkeling van datagrids zijn toepassingen waarbij grote gedistribueerde hoeveelheden data en veel gebruikers op verschillende locaties zijn betrokken. In zulke systemen is het snel raadplegen en actualiseren van informatie een groot probleem. Het snelheid van het raadplegen wordt bijvoorbeeld bevorderd door te werken met meerdere kopieën van bestanden of gegevensbanken (één kopie per locatie). Een verandering in één van de kopieën door een gegevensbank-transactie kan echter niet onmiddellijk in alle andere kopieën op de andere locaties worden verwerkt. Verwerkingssnelheid en gegevenscoherentie zijn dus tegenstrijdige doelen bij de ontwikkeling van dergelijke systemen. Met de datagrid-software *Avaki Data Grid* van het Amerikaanse bedrijf Avaki kunnen gegevens lokaal worden beheerd op de plek waar ze zich bevinden en van elders worden geraadpleegd door geautoriseerde gebruikers. Hierdoor vervalt de noodzaak van een complexe replicatie-infrastructuur.

Voorbeelden van datagrid-onderzoeksprojecten voor de gedistribueerde analyse van

grote hoeveelheden data zijn het *Grid Physics Network*¹, het *EU DataGrid-project*² en de *Particle Physics DataGrid*³. Op het gebied van genetisch onderzoek worden door de introductie van nieuwe onderzoekstechnieken, zoals DNA-microarrays, in talrijke onderzoekslaboratoria enorme hoeveelheden gegevens gegenereerd. Er is een dringende behoefte aan ontsluiting van deze bronnen voor grote onderzoeksconsortia. Datagrids zijn dan een voor de hand liggende oplossing.

Binnen de levenswetenschappen blijken onderzoeksgegevens gedistribueerd te zijn over meerdere locaties en netwerkdomeinen. Zonder gridtechnologie is het op afstand raadplegen van deze bronnen vaak niet triviaal, en het beheer ervan een tijd- en geldroevende zaak.

Verder is de verwachting dat datagrids een even grote betekenis voor de humane en sociale wetenschappen en de kunsten kunnen hebben als de rekengrids voor de natuur- en technische wetenschappen. De wetenschappers uit de genoemde gebieden die zich hierbij nog niets kunnen voorstellen, zijn getroost met het feit dat ook in de natuurwetenschappen niet iedereen rekengrids gebruikt. Een concreet voorbeeld is het Rembrandt-project, waarbij de Universiteit van Amsterdam is betrokken.

APPARATUURGRIDS Apparatuurgrids geven toegang tot dure of unieke wetenschappelijke apparatuur, zoals radiotelescopen, elektronenmicroscopen of instrumenten voor de registratie van aardbevingen. Momenteel zijn er weinig voorbeelden van apparatuurgrids. Eén daarvan is het samenwerkingsverband *Network for Earthquake Engineering Simulation*⁴ (NeesGrid). Vrijwel alle huidige apparatuurgrids zijn door universiteiten en andere onderzoeksorganisaties gerealiseerd.

In opkomst is het *sensorgrid*. Hierbij wordt een groot aantal sensoren, veelal van diverse soorten en in lokale netwerken opgenomen, gekoppeld aan rekencapaciteit en op afstand gecontroleerd en bediend. Toepassingen lopen uiteen, van controle van autoverkeer tot klimaat- en milieuonderzoek, seismologie en precisielandbouw.

TOEPASSINGSGRIDS Een toepassingsgrid geeft via het computernetwerk toegang tot bepaalde applicaties. De software die de toepassing realiseert, kan op één computer(systeem) zijn geïmplementeerd of over meerdere computers of computersystemen van één of meerdere organisaties zijn verspreid. De eerstgenoemde implementatievorm komt overeen met het *application service provider*-model, dat nu de praktijk op dit terrein domineert. Het belang van de tweede genoemde implementatievorm zal met de verspreiding van componentgebaseerde technologieën en webdiensten voor de ontwikkeling van toepassingen in de komende jaren toenemen.

ORGANISATIEGRIDS Hierbij gaat het om grids die binnen een organisatie of een deel van een organisatie worden gerealiseerd. In de context van hoger onderwijs kan dit een afdeling, instituut, faculteit, hogeschool of universiteit zijn. Afhankelijk van de omvang van het grid kan men in dit verband onderscheid maken tussen instituuts-, campus- of uni-

¹ <http://www.griphyn.org>

² <http://www.eu-datagrid.org>

³ <http://www.ppdg.net>

⁴ <http://www.neesgrid.org>

versiteitsgrids. Het computernetwerk dat hiervoor wordt gebruikt, is een intranet. Dit soort grids is momenteel in de praktijk het belangrijkste voorbeeld van gridtechnologie. Dit zal ook in de komende jaren zo blijven (zie 3.1.4).

PARTNERGRIDS Hierbij sluiten twee of meer organisaties overeenkomsten om bepaalde ICT-bronnen te delen, meestal in verband met de realisatie van bepaalde gezamenlijke projecten. Het computernetwerk dat gebruikt wordt is het internet, en de te delen ICT-bronnen bevinden zich op verschillende geografische locaties en administratieve netwerkdomeinen.

In deze groep vallen ook de grids van organisaties die geografisch verspreid zijn. Dit is bijvoorbeeld het geval bij grote universiteiten in de VS. Zo is de Universiteit van Californië over meerdere campussen verspreid.

Een Europees voorbeeld van een onderzoeksorganisatie die van een grid kan profiteren, is het *European Space Agency (ESA)*, waarvan centra en partnerorganisaties over meerdere landen zijn verspreid. Alleen al video- en audioconferencing met behulp van grid-communicatietechnologie zal voor deze organisatie een grote tijd- en kostenbesparing opleveren. Daarnaast biedt een grid voordelen voor collaboratief ontwerpen, kunnen zware rekentaken zoals mechanische analyse gedistribueerd worden uitgevoerd en grote gegevensbanken effectief worden gedeeld. ESA heeft recentelijk echter geconcludeerd dat gridtechnologie verder ontwikkeld dient te worden voordat gebruik kan worden gemaakt van de evidente voordelen.

Andere organisaties hebben een groter vertrouwen in de stand van de technologie. Zo maakt het internationale farmaciebedrijf Pfizer gebruik van de door Avaki geleverde gridsoftware om computerfaciliteiten en gegevensbanken van vestigingen op meerdere geografische locaties in een grid te integreren. Binnen Nederland kan dit model van belang zijn voor bijvoorbeeld TNO, bij fusies van universiteiten met hogescholen, of bij de realisatie van gezamenlijke opleidingen van (technische) universiteiten.

SERVICEGRIDS Een trend in de ICT-ontwikkeling van de laatste jaren is de opkomst van leveranciers (*service-providers*) van diverse ICT-diensten, zoals *web-hosting*, *content-verspreiding*, reken- en opslagcapaciteit en allerlei toepassingen. Deze leveranciers maken gebruik van schaalvoordelen om diensten tegen een lagere prijs te kunnen leveren. Voor de afnemers is dit model voordelig, aangezien deze diensten niet tot de kerntaken van de afnemer horen. Bedrijven en organisaties kunnen wezenlijke kosten besparen door het *outsourcen* van niet-essentiële onderdelen van hun ICT-infrastructuur aan dergelijke leveranciers. Binnen een universiteit of hogeschool is er overigens een onderdeel dat bij uitstek geschikt is om de rol van leverancier of bemiddelaar bij de levering van zulke diensten over te nemen: het reken- of ICT-centrum.

Een servicegrid is een volgende stap in de ontwikkeling van het *service-provider*-model. Het nieuwe hierbij is dat een veelvoud van diensten kan worden aangeboden en in grotere mate kan worden geïntegreerd. Ook hoeft een dienst niet van een bepaalde leverancier te komen. De leverancier kan binnen het servicegrid dynamisch door de afnemer worden gekozen, afhankelijk van de momentane behoefte van de afnemer, de belasting van de leverancier en de kwaliteit en prijs van de dienst die aangeboden wordt.

Dankzij grid-middleware kunnen faciliteiten die op verschillende locaties zijn

gevestigd en eigendom zijn van verschillende instituten, gecoördineerd ingezet worden voor de uitvoering van een taak van een gebruiker. Deze integratie van faciliteiten hoeft zich niet tot de reken- en opslagfaciliteiten te beperken. Programmatuur, gegevensbanken, meetinstrumenten en visualisatiefaciliteiten kunnen ook dusdanig via het netwerk worden geïntegreerd dat een gebruiker niet merkt waar deze zich bevinden. Op deze manier kunnen rond een grid *virtuele organisaties* ontstaan, bestaande uit de instituten en individuen die hun faciliteiten aan de grid beschikbaar stellen. Elke deelnemer kan daarbij beperkingen opleggen aan de voorzieningen en de plaats en het moment waarop deze ter beschikking worden gesteld.

WERELDWIJDE SERVICEGRIDS De afzonderlijke wetenschappelijke en commerciële grids kunnen worden samengevoegd tot één wereldwijd grid waar een ieder toegang toe heeft. Een dergelijk wereldwijd grid wordt als de opvolger van het World Wide Web gezien. Terwijl het World Wide Web toegang geeft tot multimediale informatie (tekst, beelden, muziek en film) op het internet, zal een wereldwijd grid toegang geven tot informatie en ICT-diensten.

Een recent voorbeeld van een project in deze richting is *PlanetLab*, dat bedoeld is als testbed voor gridtechnologie op het internet. Momenteel dragen 65 locaties van zestien landen bij aan dit project, dat wordt gecoördineerd vanuit de universiteit van Princeton. Men bestudeert het gedrag van applicaties die een beroep doen op ICT-bronnen die over de hele wereld zijn verspreid.

Een wereldwijd grid is echter nog in ontwikkeling en kan in de komende vijf, misschien tien jaar slechts een beperkte rol voor Nederlandse universiteiten en hogescholen spelen. Het ontwikkelingsstadium van wereldwijde grids is enigszins vergelijkbaar met het stadium waarin het World Wide Web zich in 1994 bevond.

Clustersystemen en intranetgebaseerde organisatiegrids

Het wereldwijde servicegrid in het groot, zoals het door visionairs wordt getekend als het delen van ICT-bronnen via internet op mondiale schaal, is er nog lang niet. Er zijn al wel voorbeelden van bedrijven en organisaties die via internet diensten van externe leveranciers gebruiken om hun ICT-infrastructuur aan te vullen.

Verreweg de meeste grids worden in eerste instantie binnen een organisatie op basis van intranets gerealiseerd. Zulke organisatiegrids (*enterprise grids*) zijn op dit moment de belangrijkste toepassing van de gridtechnologie en dit zal de komende vijf jaar zo blijven.

Intranetgebaseerde grids profiteren van de rijpe technologie en de grote ervaring op het gebied van *clustersystemen*. Een cluster is een verzameling van computers die via een snel netwerk met elkaar verbonden zijn en van systeemsoftware zijn voorzien die

⁵ Inmiddels is de prijsgedreven trend om off-the-shelf-processoren, bijvoorbeeld overwegend Intel Itanium, en andere standaardcomponenten te gebruiken, doorgebroken in de supercomputertechnologie. Het belangrijkste onderscheid tussen clusters en supercomputers ligt op dit moment in de latentietijd voor de communicatie tussen de processoren en dus in het verbindende netwerk.

het mogelijk maakt deze verzameling als een parallelle computer te gebruiken. Gewoonlijk bevinden de tot een cluster behorende computers zich in één ruimte of in één gebouw.

In de afgelopen tien jaar is grote vooruitgang geboekt in de ontwikkeling van de clustertechnologie. Clustersystemen hebben het begrip *supercomputing* een nieuwe dimensie gegeven en dramatische veranderingen op de markt van *high-performance*-systemen veroorzaakt. Dit succes is te danken aan het feit dat clustersystemen worden opgebouwd uit standaardcomponenten die in elke computerwinkel direct van de plank (*off the shelf*) verkrijgbaar zijn: pc's, werkstations of servers, het Linux-besturingssysteem, netwerkkaarten en -schakelaars. **De lage prijs van deze standaardcomponenten vertaalt zich in een prijsniveau van clustersystemen dat vele malen lager dan dat van supercomputers ligt.**⁵

Het bouwen van clustersystemen, dat aan de universiteiten is begonnen als een alternatief voor *high-performance-computing*, is nu een onderdeel van de ICT-industrie. Door hardware-leveranciers en *systeemintegrators* worden complete (*turn-key*)-clustersystemen geboden. Clustersystemen worden zowel aan universiteiten als in het bedrijfsleven gebruikt. De Rijksuniversiteit Groningen heeft bijvoorbeeld enkele clustersystemen voor wetenschappelijk rekenen; het grootste daarvan bestaat uit 128 computers.

Hoewel een clustersysteem meestal uit gewone computers is samengesteld, zoals pc's die van hun beeldschermen zijn ontdaan, wordt het in de regel gebruikt voor specifieke taken, zoals wetenschappelijke simulaties. Dit betekent dat de pc's tussentijds niet voor iets anders, bijvoorbeeld tekstverwerking, kunnen worden ingezet. De configuratie en het doel van een clustersysteem liggen dus min of meer vast en worden in de regel niet dynamisch veranderd. Dit is bijvoorbeeld het geval bij het bovengenoemde systeem van de Rijksuniversiteit Groningen. Bij de ontwikkeling van clustersoftware is echter gebleken dat de beperking van een statische configuratie geen noodzakelijke voorwaarde is. Een cluster kan ook dynamisch worden samengesteld op basis van computers die in een gegeven periode niet voor andere doelen worden gebruikt.

TOEPASSING VAN ORGANISATIEGRIDS Neem het voorbeeld van Purdue University in de VS. Deze universiteit verbindt 's nachts 2.300 computers (onderwijs-pc's en servers) tot een cluster. Overdag worden deze computers gewoon in het onderwijs gebruikt, 's nachts zijn ze beschikbaar als één clustersysteem. Omdat de integratie van de computers in een cluster dynamisch plaatsvindt, kan men in dit geval van een grid spreken. Bij Purdue wordt het dynamische clustersysteem, of grid, vooral voor onderzoek in genomics en proteomics gebruikt. Veel taken op dit terrein zijn zeer geschikt voor uitvoering op een verzameling van computers, omdat ze uit een groot aantal van elkaar onafhankelijke rekenopdrachten bestaan. Zo moeten bij Purdue ongeveer tienduizend van dergelijke opdrachten per dag worden uitgevoerd voor de bepaling van genetische sequenties. Vroeger werden hiervoor dure supercomputers ingezet. Voortaan worden de genoemde rekenopdrachten 's nachts op onderwijscomputers uitgevoerd, waardoor de supercomputers vrij komen voor andere taken. Hierdoor wordt de druk om permanent in nieuwe dure supercomputers te investeren minder groot en wordt de standaard-IT-infrastructuur beter benut. Dit levert een directe kostenbesparing op.

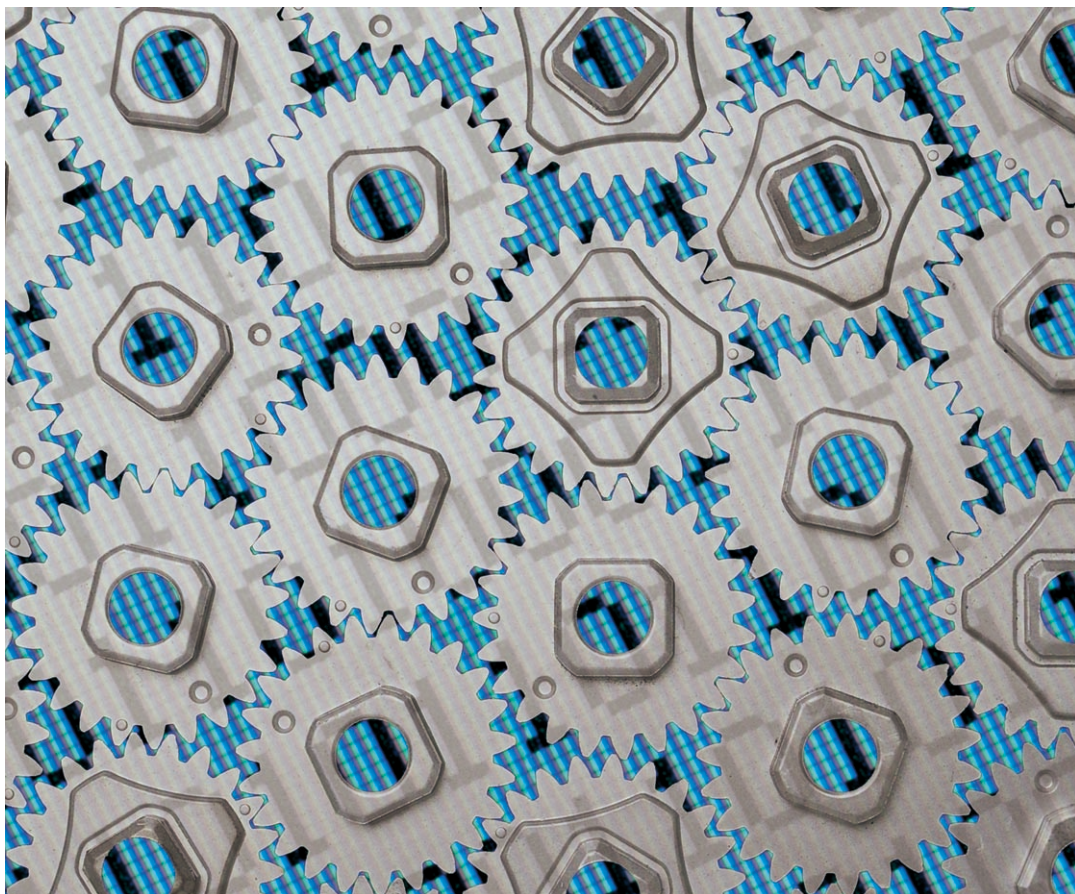
Universiteiten en hogescholen kunnen gemakkelijk hun bestaande ICT-infrastructuur aanpassen voor dit soort dynamische cluster- of grid-computing. De verbeterde systeemintegratietechnologie kan bepaalde veranderingen in het beleid teweegbrengen. Zo was een uitgangspunt bij de aanschaf van het genoemde clustersysteem van de Rijksuniversiteit Groningen dat de geavanceerde pc's waaruit het cluster is opgebouwd, na een periode van twee jaar door nieuwe pc's met betere processoren vervangen zouden worden. De oude pc's van het cluster zouden dan naar het onderwijs gaan. Er is ook een ander scenario mogelijk: het geld dat voor de vervanging beschikbaar is, wordt rechtstreeks aan onderwijs-pc's besteed; deze pc's worden 's nachts dynamisch in een cluster verbonden voor onderzoeksdoeleinden. Ook overdag kunnen onderwijs-pc's voor rekenintensieve onderzoeksprojecten worden benut. Het overgrote deel van de rekenkracht van deze computers wordt immers niet gebruikt. Taken als tekstverwerking en het lezen en versturen van e-mail, waarvoor de genoemde computers het overgrote deel van de tijd worden ingezet, eisen minder dan enkele procenten van de processortijd van een computer. De onbenutte rekentijd kan met grid-software beschikbaar worden gesteld voor rekenintensieve onderzoeksprojecten zonder het normale gebruik van deze computers te belemmeren. De rekenintensieve taken worden op de achtergrond uitgevoerd en de gebruiker merkt er niets van. Binnen een intranetgebaseerd afdelings-, instituuts- of universiteitsbreed grid kan een toepassing op een willekeurige server of pc worden uitgevoerd. Het kan dan blijken dat een kleiner aantal servers of minder krachtige pc's ook voldoende zijn, zodat op deze manier wezenlijke directe (hardware-)kosten zijn te besparen. Op de genoemde schaal zijn de veiligheidsrisico's, het tot nu toe grootste obstakel voor het gebruik van organisatieoverstijgende grids, nog te overzien en te beperken.

Commerciële gridproducten en -diensten

Hieronder wordt een aantal bedrijven en producten genoemd. De bedoeling is om de lezer een idee te geven van de breedte en omvang van het grid-productenpalet. De beperkte ruimte staat echter een volledig overzicht van de dynamiek van de markt niet toe.

Het grid-softwarepakket *Grid Engine* van SUN Microsystems is geschikt voor het verbinden van een relatief klein aantal servers, workstations of desktop-pc's van een afdeling in een cluster. Voor de gebruiker ziet het afdelingscomputersysteem dat door Grid Engine bestuurd wordt, er als een enkele computer uit. De gebruiker hoeft zijn programma's niet op zijn eigen pc of een andere specifieke computer binnen de afdeling te draaien. De Grid Engine kiest dynamisch welke computer binnen de afdeling vrij en het meest geschikt is voor de uitvoering van een programma. De Grid Engine voorkomt dat een server in het cluster overbelast wordt, terwijl andere servers niets te doen hebben. Door de verbeterde efficiëntie zal het minder snel noodzakelijk zijn om nieuwe servers aan zo'n cluster toe te voegen. Inmiddels is dit gratis door SUN Microsystems beschikbaar gestelde programma

⁶ <http://www.sun.com>



duizenden keren gedownload van de webpagina van het bedrijf.⁶ Volgens SUN Microsystems wordt de Grid Engine-software gebruikt om in totaal meer dan 100.000 processoren wereldwijd in verschillende grids te verbinden. Het pakket zou gemakkelijk en snel te installeren zijn en men zou binnen twee dagen een grid kunnen realiseren. Extra functionaliteit en mogelijkheden voor het beheren van grotere grids die uit meerdere clusters binnen een organisatie bestaan, worden geboden door de *Grid Engine Enterprise Edition* en *SUN ONE Grid Engine*. Voor het bouwen van grids uit systemen die op verschillende geografische locaties zijn gevestigd, biedt SUN het *Global Grid*-systeemprogramma.

Software voor het bouwen van grids wordt ook door andere bedrijven geboden, bijvoorbeeld Platform Computing en Enterprise United Devices (met het pakket *Grid MP*). Het systeem *DCGrid* van het bedrijf Entropia realiseert soortgelijke doelen als de Grid Engine van SUN Microsystems, maar is specifiek gericht op netwerken van Windows-gebaseerde desktop-pc's. Oracle biedt het product *Real Application Center*, dat in eerste instantie gericht is voor het draaien van zijn gegevensbanksysteem Oracle 9i op een cluster van servers, maar dat ook de uitvoering van andere applicaties op een cluster ondersteunt.

Het Amerikaanse bedrijf Avaki biedt rijpe oplossingen voor grids die over meerdere geografische locaties en internetdomeinen zijn verspreid. Het internationale farmaciebedrijf Pfizer bijvoorbeeld maakt gebruik van het pakket *Avaki Data Grid* om computerfaciliteiten en gegevensbanken van vestigingen op meerdere geografische locaties in een datagrid te integreren.

IBM is sinds enkele jaren zeer actief op het terrein van grid-computing. Het bedrijf heeft hiertoe miljarden dollars in personeel en infrastructuur geïnvesteerd. De term die IBM gebruikt voor grid-computing is (*deep*) *computing on demand*. In tegenstelling tot andere bedrijven biedt IBM niet alleen de systeemsoftware die voor het bouwen van grids nodig is: een eigen omvangrijk wereldwijd netwerk van *server-pools*, waarvan de eerste medio 2003 in gebruik is genomen, is de basis waarop computerdiensten *on demand* aan klanten kunnen worden geleverd via een VPN-verbinding binnen het internet. Klanten die veel verwerkingskracht nodig hebben, zoals olieveldontsluitingsbedrijven, de farmaceutische en de filmindustrie, financiële dienstverleners, overheidsinstanties en onderzoeksinstituten, worden aangelokt door het perspectief van directe reductie van de kosten voor infrastructuur en beheer. Aan de softwarekant is er een reeks van producten, zoals het *WebSphere*-platform, de *WebSphere-application-server*, *NAS* en *iSCSI* (specifiek voor opslag), *Storage Tank* (complete databack-up), *TotalStorage* (*off-line back-up*), *Tivoli* (beheer van grids), *eServer p-* en *xSeries* (voor aansluiting van supercomputers aan grids) en specifieke pakketten voor diverse branches, zoals gezondheidszorg, overheid, financiële dienstverlening, ruimtevaart en de autobranche.

Ontwikkeling van grid-applicaties

Met de komst van parallelle computers zijn nieuwe programmeermodellen ontstaan, zoals *dataparallel programmeren* en *message passing*. Deze modellen zijn de basis geworden van standaarden, zoals de nieuwe versies van de programmeertaal Fortran (bijvoorbeeld respectievelijk de communicatiebibliotheek *Message Passing Interface (MPI)*). De ontwikkeling van gridtoepassingen is moeilijker dan de ontwikkeling van applicaties voor gewone of geavanceerde (parallelle) computers. Grid-programmeren is gericht op de afstemming van activiteiten in heterogene dynamische omgevingen door het regelen van de interactie tussen gedistribueerde faciliteiten, diensten en gegevensbronnen. Deze nieuwe technologie noodzaakt tot de ontwikkeling van nieuwe, grid-specifieke programmeermodellen en praktijken. Hieronder volgt een overzicht van recente ontwikkelingen op dit terrein volgens Laforenza (2002).

GRID-GEORIËNTEERDE MESSAGE-PASSING- BIBLIOTHEKEN	Een <i>message-passing</i> -bibliotheek maakt de communicatie mogelijk tussen processen die op verschillende processoren kunnen worden uitgevoerd. Bij een parallelle computer zijn dit de processoren van die computer. In een gridcontext kunnen dit de processoren van computers op verschillende locaties zijn. De communicatie tussen computers op verschillende locaties is echter meerdere ordes van grootte langzamer dan de communicatie binnen een parallelle computer. Grid-georiënteerde message-passing-bibliotheeken stellen de programmeur in staat om rekening te houden met deze verschillen via aparte functies voor snelle en langzame communicatie. Er bestaan een aantal grid-georiënteerde message-passing-bibliotheeken: <i>MagPIe</i> ,
---	---

MPICH-G2, MPI_Connect, MetaMPICH, PACX-MPI, PVMPI. Hoewel deze benadering op een behoorlijk laag programmeerniveau is gepositioneerd, lijkt hij de enige plausibele optie voor de ontwikkeling van efficiënte gridtoepassingen in de nabije toekomst.

NETWORK-ENABLED SERVER Hierbij gaat het om het gebruik op afstand van programmabibliotheken en rekenfaciliteiten. Op dit moment zijn voorbeelden te vinden op het terrein van numerieke berekeningen (*Ninf, NetSolve*). De oplossing van een stelsel lineaire vergelijkingen is bijvoorbeeld een veel voorkomend probleem bij wetenschappelijke simulaties en technische berekeningen. De gebruiker kan vanuit een (client)programma, dat geschreven is in een traditionele programmeertaal als Fortran of C, een functie oproepen voor de oplossing van zo'n stelsel. Via het globale netwerk wordt de functieoproep, samen met parameters zoals een matrix van coëfficiënten, doorgegeven aan een softwarebibliotheek- en rekenserver. De server voert de berekeningen uit en retourneert de resultaten.

Aan de basis van deze benadering liggen de traditionele concepten *remote procedure-call* en *client/server*. Een middleware-laag zorgt voor de afhandeling van de functieoproep, inclusief het vinden van een beschikbare server die de gevraagde dienst kan verlenen. De middleware zorgt tevens voor authenticatie, autorisatie en *billing*.

COMPONENT-GEBASEERDE TECHNOLOGIEËN Een component is een uitvoerbaar deelprogramma dat gedefinieerd is door zijn functie en het protocol dat het gebruikt om met andere componenten te communiceren. De grote potentie van deze benadering voor de beheersing van de complexiteit van grote toepassingen en voor het hergebruik van software is door het bedrijfsleven in voldoende mate herkend. Dat heeft geleid tot de ontwikkeling van systemen als *Common Object Request Object Architecture (CORBA)*, *Distributed Component Object Model (DCOM)*, *Enterprise JavaBeans (EJB)*, en *.NET*.

De componentgebaseerde manier waarop huidige complexe toepassingen worden gebouwd, is zeer geschikt voor gebruik in gridtoepassingen. De componenten van een toepassing kunnen op verschillende computers in een grid worden uitgevoerd. Hierbij gaat het om componenten waarvan de uitvoering minimaal enkele minuten vergt. Voor zulke componenten speelt de communicatietijd in het grid geen grote rol en lijdt de efficiëntie er niet onder. De verwachting is dat met de toenemende beschikbaarheid van grid-middleware de componentgebaseerde technologieën op middellange termijn uitgroeien tot dé manier om gridtoepassingen te ontwikkelen.

WEB-DIENSTEN Het World Wide Web is ontstaan als een systeem voor snelle en gebruikersvriendelijke uitwisseling van informatie. In eerste instantie ging het daarbij om teksten: HTTP is een protocol voor tekstuitwisseling. De markeertaal HTML biedt slechts faciliteiten voor het opmaken van tekst en de integratie daarin van beeld en geluid. De taal XML gaat een stap verder door faciliteiten te bieden voor semantische markering van informatie. Daardoor is het mogelijk bedrijfsprocessen van twee (of meerdere) organisaties, bijvoorbeeld afnemer en toeleverancier, op elkaar af te stemmen.

Een volgende stap in de ontwikkeling van het web is om met diensten op een soortgelijke manier om te gaan. In plaats van informatie worden diensten aangeboden en afgenomen. Dan gaan we van een tekst- en informatie-web naar een *diensten-web*.

De diensten kunnen zeer divers zijn: specifieke softwareapplicaties, toegang tot gegevensbanken, dataopslag- of rekenfaciliteiten.

Net zoals bij het *net-enabled* serverconcept worden bij het concept van webdiensten softwarefuncties vanuit een applicatie aangeroepen. Deze functies implementeren de gevraagde diensten en kunnen worden uitgevoerd op servers van de eigen organisatie of van een externe leverancier. De diensten maken gebruik van verzamelingen van standaarden, die ontwikkelaars in staat stellen om gedistribueerde toepassingen te implementeren met gebruik van uiteenlopende gereedschappen van verschillende leveranciers. De meest bekende protocollen en standaarden zijn: *XML*; *Universal Description, Discovery and Integration (UDDI)*; *Simple Object Access Protocol (SOAP)* en *Web Service Description Language (WSDL)*.

Omdat HTML slechts op de opmaak van tekst is gericht, is deze taal relatief simpel, hetgeen tot zijn snelle acceptatie heeft geleid. ICT-diensten kunnen daarentegen zeer divers zijn. Dit kan ertoe leiden dat de middelen voor de beschrijving van zulke diensten dermate omvangrijk worden dat dit hun acceptatie in de praktijk belemmert. Of door het concept van webdiensten het web dezelfde rol kan spelen bij het gebruik van gedistribueerde ICT-bronnen als bij het verkrijgen van informatie, moet dus nog uit de praktijk van de komende jaren blijken.

SCRIPTING PROGRAMMEERTALEN Traditionele programmeertalen als Fortran, C of Java zijn geschikt om monolitische programma's te schrijven. Ze bieden echter onvoldoende middelen om programma's te integreren door bijvoorbeeld in de uitvoer van een programma naar bepaalde patronen te zoeken en het resultaat als invoer te sturen naar een ander programma, dat mogelijk op een andere computer wordt uitgevoerd. Een aantal *scripting* programmeertalen, zoals *Perl*, *Python*, *Tcl/Tk*, *Java-scripts* en *Unix-shells*, is uit deze behoefte ontstaan.

Deze benadering biedt een werkbare manier om snel gedistribueerde toepassingen te bouwen en wordt zeer breed in de praktijk toegepast. De verbinding van webserver, die informatie aan webgebruikers beschikbaar stellen of informatie ontvangen, en gegevensbankserver wordt bijvoorbeeld in de regel gerealiseerd door een script-programma dat in zo'n *scripting* programmeertaal is geschreven.

PROBLEEMOPLOS- OMGEVINGEN *PSE's (Problem Solving Environments)* zijn systemen die alle computerfaciliteiten beschikbaar stellen die nodig zijn voor de oplossing van een gegeven klasse van problemen. Het begrip stamt van vóór het gridtijdperk, toen dergelijke omgevingen ontwikkeld werden voor specifieke computersimulaties in bijvoorbeeld de stromingsleer. Het doel is de productiviteit van onderzoekers te verhogen door ondersteuning te bieden voor probleemomschrijving, keuze van een oplosmethode, uitvoeren van simulaties en analyse van de resultaten in een geïntegreerde omgeving. De onderzoeker kan zich concentreren op het oplossen van zijn probleem en wordt niet afgeleid door de noodzaak gereedschappen te ontwikkelen. Een gridcontext voegt aan dit doel nog een dimensie toe. De faciliteiten die in een PSE worden ingezet, kunnen over verschillende heterogene computersystemen zijn gedistribueerd, zonder dit een zorg voor de eindgebruiker hoeft te zijn. Systemen die het bouwen van gridgerichte PSE's ondersteunen, zijn bijvoorbeeld *WebFlow* en *Commodity Grid (CoG)*.

FRAMEWORKS *Frameworks* zijn geïntegreerde verzamelingen van software-tools die de ontwikkeling van applicaties vergemakkelijken. Twee voorbeelden zijn Cactus en Meta-Chaos. *Cactus* wordt onder meer gekenmerkt door een modulaire opbouw, die de parallelle uitvoering op diverse platformen en de samenwerking van verschillende groepen bevordert. Er zijn modules voor diverse standaardtaken, zoals parallelle I/O, data-distributie of *checkpointing*, en modules voor specifieke toepassingen, zoals de simulatie van verbrandingsprocessen. Nieuwe toepassingen worden in een soort objectgeoriënteerde metaprogrammeertaal geschreven, door aan te geven hoe uitvoerbare programma's onderling gegevens uitwisselen. *Meta-Chaos* is gericht op de uitwisseling van gegevens tussen dataparallelle bibliotheken.

PORTALS Voorbeelden van een *portal* in een generieke internetcontext zijn *Yahoo*, *Alta Vista*, *AOL*, *Lycos* en *Infoseek*. In de context van wetenschappelijk rekenen geeft een portal toegang tot geïntegreerde diensten die wetenschappelijke simulaties mogelijk maken. *HotPage* is bijvoorbeeld de *grid-computing-portal* van NPACI (NSF's Partnerships for Advanced Computational Infrastructure). Door dit portal kan een gebruiker van de grid-computingfaciliteiten van NPACI een overzicht krijgen van de op een gegeven moment beschikbare bronnen, en bepalen met welke faciliteiten en op welke manier zijn taak wordt uitgevoerd.

COMPILATIESYSTEMEN VOOR HET GRID Uit de praktijk is bekend dat gecompileerde programma's het snelst zijn: dat wil zeggen programma's die voor hun uitvoering in machinecode zijn omgezet en waarbij ook de toewijzing van geheugen heeft plaatsgevonden. Bij de ontwikkeling van parallelle computers met gedistribueerd geheugen is gebleken dat de compilatie van programma's op zulke systemen het moeilijkste probleem is. Dit heeft tot de ontwikkeling van alternatieve oplossingen geleid, zoals de message-passing-bibliotheken: de programmeur geeft aan hoe de gegevens over de geheugens van de afzonderlijke processoren worden verdeeld en zorgt ervoor dat ze zonnodig van het ene naar het andere geheugen verplaatst worden. In een gridcontext lijkt de compilatie een nog moeilijker probleem. In een parallelle computer is het aantal processoren en hun verbindingen van tevoren bekend. De configuratie van de in een grid verbonden computers verandert echter permanent. Bij de uitvoering van een gedistribueerd programma dient dynamisch rekening te worden gehouden met de veranderingen in de gridconfiguratie. De verwachting is dat deze benadering pas op de lange termijn vruchten kan afwerpen.

Voorbeelden van grote gridprojecten

Alleen al in de VS zijn sinds medio jaren negentig enkele honderden miljoenen dollars door de overheid beschikbaar gesteld voor gridonderzoek. Ook elders in de wereld zijn er enkele tientallen gridonderzoeksprojecten gaande. Toepassingen zijn onder meer te vinden op de terreinen van hoge-energiefysica, modelleren van het klimaat, genetisch onderzoek, simulaties van aardbevingen, sterrenkunde en hersenenonderzoek.

Een recent rapport van de National Science Foundation in de VS, getiteld

Revolutionizing Science and Engineering Through Cyberinfrastructure, pleit voor nieuwe financiering in de omvang van één miljard dollar per jaar (!) om van grid-computing een regulier onderzoeksmiddel te maken.

TERAGRID Een van de grootste gridonderzoeksprojecten op dit moment is *TeraGrid*⁷ van de *National Science Foundation*. Dit meerjarig project heeft als doel het bouwen van 's werelds snelste en meest complexe gedistribueerde computerinfrastructuur voor wetenschappelijk onderzoek. Het hart van dit systeem wordt gevormd door krachtige parallelle computers op vijf geografische locaties. De totale verwerkingskracht zal 20 teraflops bedragen. Naast grote verwerkingskracht zal het systeem een enorme opslagcapaciteit beschikbaar stellen: bijna 1 petabyte (10^{15}). Visualisatiefaciliteiten en gegevensbanken behoren ook tot deze infrastructuur. De genoemde computers en faciliteiten zullen door een speciaal snel netwerk van 40 gigabit per seconde worden verbonden.

GRIDPROJECTEN BIJ CERN Negen grote Europese rekencentra werken samen in het kader van het door de EU gesubsidieerde *DataGrid*-project. De deelnemers willen hun reken-, opslag- en netwerkcapaciteiten op een gecoördineerde manier laten samenwerken voor gedistribueerde analyse van gedistribueerd opgeslagen data (honderden petabytes), die verkregen zijn met een nieuwe deeltjesversneller (*Large Hadron Collider*) van het Europees centrum voor nucleair onderzoek CERN.⁸ Dit geheel moet in 2005 operationeel zijn. Bij de gegevensanalyse van de deeltjesversneller zullen maar liefst vijfduizend wetenschappers van 150 universiteiten betrokken zijn. Een ander voorbeeld van grootschalige en gedistribueerde data-opslag in dit kader is het project *Globally Interconnected Object Databases (GIOD)*, dat gericht is op de data die gegenereerd zullen worden bij CERN.

GRID PHYSICS NETWORK Een van de meest ambitieuze gridprojecten van dit moment is het *Grid Physics Network (GriPhyN)*. Dit samenwerkingsverband wordt gefinancierd door de Amerikaanse National Science Foundation. De verwachting is dat de cumulatieve rekenkracht van GriPhyN meer dan 120 triljoen operaties per seconde gaat bedragen. Om dit te bereiken zullen wereldwijd duizenden computers, niet uitsluitend supercomputers, ingezet worden.

RADIOASTRONOMIE Radiotelescopen maken simultaan gebruik van meerdere antennes. Door de met deze antennes gewonnen signalen gezamenlijk te analyseren, krijgt men een grotere resolutie aan de hemel. Hoe meer antennes men gebruikt en hoe verder uit elkaar deze antennes staan, hoe groter het oplossingsvermogen. Tot nu toe was het in de radioastronomie gebruikelijk *antenna-arrays* te bouwen. Voorbeelden daarvan kunnen in Dwingeloo en Westerbork worden gezien. Door antennes in verschillende landen met elkaar te verbinden, kan men een nog grotere resolutie bereiken. Vereist hiervoor is een snel datanetwerk en veel rekenkracht en opslagcapaciteit. Radioastronomen maken derhalve plannen om alle radiotelescopen in Europa te

⁷ <http://www.teragrid.org>

⁸ CERN stond overigens aan de wieg van de ontwikkeling van het World Wide Web.

verbinden met de krachtige processor bij *JIVE (Joint Institute for VLBI in Europe)*, dat gehuisvest is in Dwingeloo. Met SURFnet in Nederland en Géant in Europa worden in 2003 Gb/s-dataverbindingen aangelegd van de telescopen naar Dwingeloo. Vanaf 2004 hebben de astronomen daarmee een radiotelescoop zo groot als heel Europa. De volgende ontwikkeling in de radiosterrenkunde richt zich op een ambitieus project, *LOFAR*, dat voor een groot deel in Nederland bedacht en gepland wordt. Deze telescoop bestaat uit meer dan tienduizend eenvoudige radioantennes, verspreid over een gebied met een diameter van ongeveer 350 km. Deze verzameling van antennes is feitelijk een groot wide-area-sensornetwerk, waarbij de benodigde netwerkcapaciteit en rekenvermogen te vergelijken zijn met die van het *Distributed Tera-GRID Facility (DTG)* in de VS (zie tabel 1).

Tabel 1
Vergelijking van de radio-
astronomie-projecten DFT
en LOFAR

Project	Begin	Rekenkracht [Tflops]	Opslag [Tbyte]	Bandbreedte [Gb/s]
DFT	2001	12	450	40
LOFAR	2003	40	600	160n

n = aantal kleuren op het LOFAR-interne netwerk voor datatransport

Binnen het LOFAR-project wil men dan ook verder gaan dan slechts antennes met elkaar en met rekencapaciteit verbinden. De voorgenomen rekenkracht van het systeem wordt groot genoeg om de individuele signalen tot achttmaal toe te kopiëren en er simultaan in software een achttal onafhankelijke telescopen van te maken. Hierbij begint het grid een rol te spelen. Men wil met gridtechnologie's werelds eerste *multi-user, multi-tasking, on-line* software-telescoop bouwen, die vanuit operationele centra in meerdere landen kan worden aangestuurd. De eerste drie centra zijn gepland bij de Rijksuniversiteit Groningen, bij het MIT in Cambridge in de VS en bij de Universiteit van Sydney in Australië. Verwacht wordt dat deze centra de gegevens naar gebruikers bij universiteiten overal in de wereld veelal in *quasi-real-time* zullen doorsturen. Uitgezocht wordt zelfs hoe één van de telescopen voor het algemene publiek via het internet ter beschikking kan worden gesteld. Deze gedachtengang leidt al gauw tot ideeën voor het koppelen van andere soorten sensoren aan de netwerk-infrastructuur van LOFAR. Gepland wordt om seismische sensoren aan te sluiten, om zodoende driedimensionale beelden met ongekend scherpe details van de aardgasreservoirs onder Noord-Nederland te verkrijgen. Tevens wordt uitgezocht hoe milieusensoren voor experimenten in de precisie-landbouw aangekoppeld kunnen worden. Om optimaal gebruik van deze sensor-arrays te realiseren, zijn uiteraard speciale operationele centra op afstand met gridtechnologieën nodig.

BIOMEDICAL INFORMATICS RESEARCH NETWORK Dit grid ter bevordering van hersenenonderzoek is met hulp van de National Institutes of Health van de VS in 2001 gestart. Een deelproject, met de titel *Brain Morphometry BIRN*, gaat bijvoorbeeld over het beschikbaar stellen en verwerken van MRI-data voor onderzoek naar de ziekte van Alzheimer. Onderzoekers van verschillende deelnemende universiteiten en ziekenhuizen kunnen de gedistribueerde gegevens via een datagrid raadplegen.

Voorbeelden van andere gridprojecten zijn *FusionGrid*, het *Science Grid* van de Department of Energy van de VS, het *Network for Earthquake Engineering Simulation*, de *International Virtual Data Grid Laboratory* en het *NSF Middleware Initiative*.

NEDERLANDSE GRID-ONDERZOEKSPROJECTEN Aan Nederlandse universiteiten zijn verschillende groepen betrokken bij grid-activiteiten. Binnen de informatica-onderzoekschool ASCI wordt gewerkt aan een *testbed* voor gedistribueerd rekenen, *DAS* genoemd, dat deel uitmaakt van het Europese testbed *Egrid*. Met hulp van de stichting Nationale Computer Faciliteiten (NCF) van NWO wordt dit systeem naar een Nederlands gridtestbed uitgebreid. Zo worden aan deze verzameling computers systemen voor rekenen en dataverwerking door Nikhef en Astron (*Dutchgrid*) toegevoegd. Verder is NCF betrokken bij het *NEG* (*Northern European Grid*), een onderdeel van de *EGEE*-aanvraag bij de EU (opvolger van het Europese Datagrid-initiatief). Nederlandse informatici nemen ook deel aan het Europese *GridLab*-project. De middleware die uit dit project moet voortvloeien, dient als tussenlaag te fungeren tussen een data- en rekenintensieve astrofysische applicatie en de Globus-grid-middleware. Aan de applicatiekant zijn er binnen Nederland activiteiten bij het Nikhef en het KNMI, gericht op het Europese DataGrid-project voor snelle verwerking van grote datastromen. Wat het gebruik van grids voor e-science betreft, verwijzen we naar de ICES/KIS-projecten *Virtueel Laboratorium voor E-science ontwikkeling* en *Virtueel Laboratorium Nederland* van FOM en andere partijen. Tenslotte is de deelname van stichting Astron van NWO aan het Europese LOFAR-project te noemen.

Projecten met een goed doel en betrokkenheid van particulieren

Op dit moment is er een aantal projecten waarbij burgers een deel van de verwerkingskracht van hun pc's aan een organisatie met een goed doel ter beschikking kunnen stellen. Bij het project *SETI@home* (*Search for Extraterrestrial Intelligence*), dat circa vijf jaar geleden is begonnen, wordt bijvoorbeeld naar sporen van buitenaards leven in gegevens uit radiotelescopen gezocht. Bij een ander project, *Find-a-Drug*⁹, dat in 2002 van start is gegaan, gaat het om de ontwikkeling van nieuwe geneesmiddelen tegen ziektes als kanker, pest, multiple sclerose, SARS en AIDS. Na de terroristische aanslagen van 11 september 2001 werd in een soortgelijk project naar een vaccin tegen het pokkenvirus gezocht. In een Japans project wordt gezocht naar overeenkomsten en specifieke patronen in de genetische informatie van mensen en de relatie ervan met bepaalde ziektes.

Vrijwel altijd gaat het om wetenschappelijke problemen die uit een groot aantal onafhankelijke en identieke (op de te verwerken gegevens na) deelproblemen bestaan. Een voorbeeld is het onderzoek naar de potentie van vele verschillende molecuulstructuren om als geneesmiddel tegen een bepaalde ziekte te worden gebruikt: voor elke molecuulstructuur worden de mogelijke interacties met een eiwit van de ziekteveroorzaker onderzocht.

⁹ <http://www.find-a-drug.org/>

Iedereen die een aan het internet aangesloten computer bezit, kan deelnemen aan zo'n project. Een deelnemer kan van een server van het project een kopie downloaden van een programma dat de bewerkingen uitvoert, en een deel van de te verwerken gegevens, bijvoorbeeld een aantal molecuulstructuren. Als het programma eenmaal op de pc van de deelnemer is geïnstalleerd, is geen verdere interactie met de pc-eigenaar nodig. De pc voert het programma op de achtergrond uit, waarbij het normale gebruik van de computer niet wordt belemmerd. Het programma maakt gebruik van de rekentijd die overblijft; in de regel is dat het overgrote deel. Bij een typische pc-taak als tekstverwerking wordt minder dan 1% van de processortijd gebruikt en blijft de resterende 99% onbenut.

Het programma stuurt de resultaten terug naar een server van het project en haalt nieuwe te verwerken gegevens op. Dit gaat door zolang de pc-eigenaar aan het betreffende project deel wil nemen. Het Nederlandse bedrijf Xolo X combineert dit verwerkingsmodel met een dienst voor het uitwisselen van muziek: gebruikers kunnen gratis muziek downloaden, maar moeten in ruil daarvoor de verwerkingskracht van hun computer aan Xolo X ter beschikking stellen. Het bedrijf verkoopt vervolgens deze verwerkingskracht aan onderzoeksinstituten.

Op dit moment zijn bijvoorbeeld meer dan achtduizend pc's van circa zestig landen bij het Find-a-Drug-project geregistreerd. Ze zijn voor dit project ingezet ter waarde van meer dan 250 jaar rekentijd. Op deze manier zijn inmiddels meer dan een half miljard moleculen onderzocht. Bij een soortgelijk project in Japan heeft men twaalfduizend pc's ingezet om in 130 dagen een wetenschappelijk probleem door te rekenen dat op een enkele computer meer dan zeshonderd jaar aan rekentijd had gekost.

KANTTEKENINGEN Strikt genomen gaat het bij zulke projecten niet echt om grid-computing in de zin van verwerkingskracht 'uit de muur'. Het gaat om gedistribueerde gegevensverwerking met internet als communicatiemiddel. Het lijkt er echter op dat de begrippen *distributed computing* en internet-computing, die hier zeker aan de orde zijn, inmiddels zijn opgegaan in het begrip grid-computing, dat bijna alles overkoepelt wat met het internet te maken heeft.

Het type rekenkracht dat men bij dit soort projecten 'uit het stopcontact' krijgt, heeft beperkingen. Het doel van zo'n project is altijd de oplossing van een concreet probleem, maar niet elk wetenschappelijk probleem kan worden aangepakt door het te verdelen in vele onafhankelijk van elkaar door te rekenen deelproblemen. Ongeacht de beperkingen denkt het Japanse bedrijf NTT Data Corp op termijn de organisatie van dergelijke projecten (inclusief de project-servers) als product te kunnen verkopen. Doelgroep zijn in eerste instantie onderzoeksinstituten en universiteiten. Deze onderzoeksinstituten hoeven dan immers geen dure supercomputers voor de oplossing van wetenschappelijke problemen te kopen.

Deze specifieke vorm van grid-computing is slechts beperkt inzetbaar en biedt een relatief smalle basis voor universitair beleid op dit terrein. Toch is dit model niet te onderschatten voor zowel het vinden van aanvullende ICT-bronnen voor academisch onderzoek als voor de uitstraling van dit onderzoek naar de rest van de maatschappij. Het gaat bij de lopende projecten immers om maatschappelijk zeer relevante onderwerpen die het grote publiek aanspreken. Bij het project voor het pokkenvirus-vaccin zouden bijvoorbeeld meer dan twee miljoen computers aangesloten zijn. Zulke pro-

jecten zijn dus goede ambassadeurs van het academisch onderzoek bij de overheid en de bevolking.

Ook in Nederland laten zich projecten formuleren die zijn afgeleid van maatschappelijk relevante onderwerpen – bijvoorbeeld zorg, milieu, veiligheid, of waterbeheer – en waarvoor zeker een groot draagvlak bij de bevolking te vinden is. Relatief kleine stimuleringsprogramma's (vanuit OCW, SURF of de universiteiten zelf) kunnen een groot rendement opleveren in de vorm van maatschappelijke herkenning, inzicht en betrokkenheid van de gewone burger bij wat in de 'ivoren toren' gebeurt. Ook de ICT-bronnen (de beschikking over de rekenkracht van mogelijk vele duizenden pc's van particulieren) die men op deze manier gratis kan krijgen, zijn natuurlijk niet te onderschatten.

Gridorganisaties en -standaarden

De gridactiviteiten van de laatste acht jaar hebben geleid tot de ontwikkeling van specifieke protocollen, standaarden, diensten en tools die zich kenmerken door openheid en interoperabiliteit. Deze twee aspecten zijn onontbeerlijk voor de realisatie van de hoofddoelstellingen van het gridconcept: het delen van ICT-bronnen en het dynamisch vormen van virtuele organisaties. De protocollen en standaarden zijn voor belangstellenden beschikbaar bij de twee grote consortia die het gridconcept ontwikkelen en promoten: het Globus-project en Global Grid Forum.

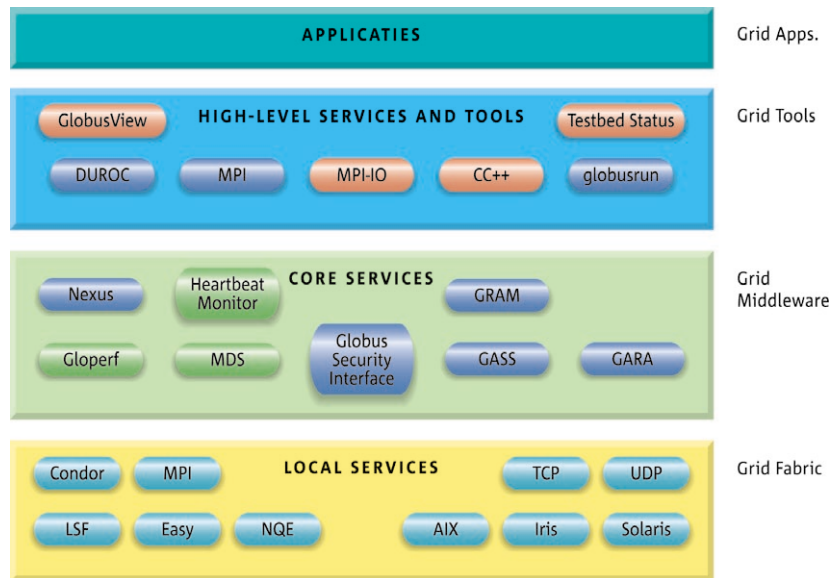
GLOBUS *Globus¹⁰ is in 1995 begonnen als een door de Amerikaanse overheid gefinancierd fundamenteel onderzoeks- en ontwikkelingsproject in informatietechnologie. De deelnemers waren enkele universiteiten en onderzoekslaboratoria in de VS. Inmiddels zijn er veel deelnemers bij gekomen, waaronder internationale en industriële partners, en krijgt dit project ook steun van bedrijven als IBM en Microsoft. In het begin was het vooral gericht op de ontwikkeling van basistechnologieën die nodig zijn om grids voor wetenschappelijk-technisch rekenen te bouwen. Tegenwoordig zijn de doelen en de resultaten van het project even belangrijk voor toepassingen in het bedrijfsleven. De *Globus Toolkit*, die in het kader van het project is ontwikkeld, is een open-source-verzameling van diensten en bibliotheken die het bouwen van grids en grid-toepassingen mogelijk maakt. Onder meer worden aspecten ondersteund als online delen, ontsluiting en management van ICT-bronnen en -gegevens, veiligheid, communicatie en portabiliteit. De Globus Toolkit is inmiddels een de-factostandaard voor het bouwen van gridsystemen geworden voor zowel academisch onderzoek als commerciële toepassingen. Het is een voorbeeld van succesvolle technologietransfer. Net als het open-source-besturingssysteem Linux wordt de Globus Toolkit ondersteund en verder ontwikkeld door een grote en groeiende groep van vrijwilligers en beroepsprogrammeurs. De toolkit is gratis verkrijgbaar en gebruikers kunnen de software naar behoefte aanpassen. Grote software-bedrijven (zoals IBM en Platform Computing) ondersteunen de Globus Toolkit. De combinatie van open source, open*

¹⁰ <http://www.globus.org>

¹¹ <http://www.gridforum.org>

¹² Foster et al. 2002

Figuur 36
Grid-architectuur.
Bron: Globus project



standaarden en de continuïteit die door het commitment van grote bedrijven is gewaarborgd, is een succesrecept voor brede acceptatie, zoals de ervaringen met het besturingssysteem Linux ons leren.

GLOBAL GRID FORUM De Global Grid Forum¹¹ wordt gevormd door individuele onderzoekers en beroeps-programmeurs die op het terrein van gridtechnologieën werken. Terwijl het Globus-project op ontwikkeling en implementatie van software is gericht, is het hoofddoel van de Global Grid Forum de ontwikkeling van standaarden, documentatie van *best practices* en richtlijnen voor implementatie. Tevens richt het forum zich op de ontwikkeling van een brede geïntegreerde grid-architectuur, die onderzoek, ontwikkeling en gebruik van nieuwe grids kan bevorderen.

OPEN GRID SERVICES ARCHITECTURE Gelet op hun ervaring en expertise op het gebied van internet-technologie en parallel en gedistribueerd rekenen, spelen de universiteiten een belangrijke rol bij het definiëren van open standaarden voor grid-computing.

De *Open Grid Services Architecture (OGSA)*, die door de Global Grid Forum wordt ontwikkeld, is een nieuwe stap naar de integratie van de gridtechnologieën die binnen de Globus Toolkit zijn ontwikkeld, in de opkomende webdienst-standaarden.¹² Deze architectuur gaat uit van een uniforme representatie van verwerkings- en opslagbronnen, netwerken, gegevensbanken, applicatieprogramma's en dergelijke. Deze worden als diensten (services) gezien die via het netwerk gebruikt kunnen worden door de uitwisseling van boodschappen. (Andere systemen, zoals *SOS* en *Legion*, gebruiken hiervoor de term 'object'.)

OGSA is gebaseerd op de integratie van concepten en tools uit de grid-computing, zoals gedefinieerd in het Globus-project, en webservice-standaarden als XML, WSDL en SOAP. Naar verwachting zullen eind 2003 alle belangrijke grid-infrastructuren op OGSA gebaseerd zijn of daaraan voldoen. IBM en Oracle hebben hun commitment uitgesproken hun producten conform de eisen van OGSA te implementeren. Tot de

bedrijven die OGSA willen gebruiken, behoren Avaki, Cray, Entropia, Hewlett-Packard, Platform Computing, Silicon Graphics, Inc en Veridian.

Conclusies en aanbevelingen

In de komende vijf jaar kunnen grids een belangrijke rol gaan spelen voor de ICT-infrastructuur van Nederlandse universiteiten en hogescholen. Een grid kan op basis van de bestaande ICT-infrastructuur binnen een afdeling, instituut, faculteit of universiteit of hogeschool worden gerealiseerd met behulp van specifieke middleware die al dan niet gratis te verkrijgen is.

Grids kunnen worden ingezet voor een efficiënter gebruik van de aanwezige ICT-infrastructuur, voor de reductie van de directe ICT-investeringen en beheerskosten, voor toegang tot hoge computercapaciteit, voor gebruik van geïntegreerde gedistribueerde gegevensbanken en toepassingen, voor interactief samenwerken op afstand en voor experimenten met apparatuur op afstand.

Bij een universiteit is er voldoende ICT-kennis om grids te laten realiseren, door eigen personeel of door een extern bedrijf. De invoering van gridtechnologie in de organisatie stelt echter eisen aan beleid, organisatie en management. De hoofdproblemen hierbij zijn de veiligheidsbedenkingen van het betrokken beheerspersoneel en de verweving van de ICT-budgetten van verschillende organisatieonderdelen door het gebruik van een grid.

De grootste kansen in het hoger onderwijs en onderzoek in de komende vijf jaar bieden computerfaciliteitengrids in een organisatie(onderdeel), die reken- en opslag-faciliteiten in een administratief netwerkdomein via een intranet integreren. Dergelijke grids laten zich zonder veel technologische en veiligheidsrisico's realiseren in een afdeling of instituut met de inmiddels rijpe gridtechnologie, en kunnen wezenlijke besparingen in de directe hardware- en beheerskosten opleveren. Wie deze kansen links laat liggen, zal teveel voor de nodige ICT-infrastructuur betalen en daarvan suboptimaal gebruik maken.

De grootste ervaring is opgedaan met rekengrids. De meeste huidige grid-onderzoeksprojecten en commerciële producten zijn op deze rekengrids gericht. Ze zijn in eerste instantie belangrijk voor wetenschappelijke en technische simulaties waarmee veel berekeningen gemoeid zijn. Voordat een universiteitsbestuurder zijn toestemming geeft te investeren in een nieuwe dure high-performance-computer voor wetenschappelijk onderzoek, zou hij eerst moeten laten uitzoeken of de gestelde doelen kunnen worden gerealiseerd met een rekengrid op basis van de aanwezige ICT-infrastructuur. Bijvoorbeeld met pools van onderwijs-pc's die 's nachts als een cluster voor parallel gedistribueerd rekenen gebruikt worden.

¹³ <http://www.gridtoday.com>

De verwachting is dat in de komende periode datagrids ook in de levenswetenschappen een zekere rol gaan spelen, in het bijzonder de genetica en de bio-informatica, en tevens in de humane en sociale wetenschappen en de kunsten.

Wat de schaal betreft waarop de integratie en het delen van ICT-bronnen plaatsvinden: de komende periode zullen de organisatiegrids, die binnen een afdeling, instituut, faculteit, hogeschool of universiteit op basis van een intranet worden gerealiseerd, het talrijkst zijn. Het aantal partnergrids, waarbij twee of meer organisaties ICT-bronnen delen, meestal om gezamenlijke projecten of opleidingen te realiseren, zal toenemen. Een wereldwijd servicegrid zal ook in de komende periode nog een toekomstvisie blijven.

Grid-onderzoeksprojecten waarbij vele duizenden burgers een deel van de rekenkracht van hun computers via internet ter beschikking stellen voor de oplossing van wetenschappelijke vraagstellingen, hebben een grote uitstraling naar de rest van de maatschappij. Ook in Nederland laten zich projecten formuleren die van maatschappelijk zeer relevante onderwerpen zijn afgeleid – denk aan zorg, milieu, veiligheid en waterbeheer – en waarvoor een groot draagvlak bij de bevolking te vinden is. Relatief kleine stimuleringsprogramma's (vanuit OCW, SURF of de universiteiten zelf) op dit terrein kunnen een groot rendement opleveren.

- REFERENTIES** D. Laforenza, 'Grid programming: some indications where we are headed', in: *Parallel Computing* 28 (2002), 1733-1752
 H. van der Pluijm, 'Verwerkingskracht uit de kraan', in: *LAN Internetworking Magazine*, 31-05-2003
 I. Foster, C. Kasselmann, J.M. Nick, S. Tuecke, 'Grid services for distributed system integration', in: *IEEE Computer*, juni 2002, 37-46.
 Meerdere berichten in *Grid Today*¹³